# Learning Conditional Probabilities for Dynamic Influence Structures in Medical Decision Models

**Cungen Cao, Tze-Yun Leong**
Medical Computing Laboratory
Department of Information Systems and Computer Science
National University of Singapore, Singapore 119260

*Based on the DynaMoL (a Dynamic decision Modeling Language) framework, we examine the critical issues in automated learning of numerical parameters from large medical databases; present a Bayesian method for learning conditional probabilities from data; analyze how to elicit prior probabilities from the domain expert; and examine several important issues on pre-processing raw data for application in dynamic decision modeling.*

## INTRODUCTION

In this work, we examine the critical issues in automated learning of probabilistic parameters from large medical database. Our discussions are based on the DynaMoL framework and a case study in the follow-up of patients who have undergone colorectal cancer surgery. We present a Bayesian method for learning conditional probabilities from data for influence views, a key representation structure in DynaMoL.

## METHODS

### Data Preprocessing

The data in a database may not be appropriate for machine learning directly. A few issues need to be addressed before learning algorithms can be applied. We propose a rule-based method for filling in missing data, and deriving values for new variables in a decision model. For the data sparseness problem, we propose data abstraction and model abstraction as a solution.

### Machine Learning

In Bayesian learning literature, the Dirichlet or Beta distributions are commonly used as prior distributions for model parameters. This strategy simplifies the learning process, because the inherent mathematical properties allow these distributions to flexibly assume various patterns in terms of locations, dispersions, shapes, *etc.*

### Exponents Assessment

A key issue here is how to assess exponents for Dirichlet distributions. For dynamic decision making, we extend the notion "equivalent information size" to a *dynamic* version of the "equivalent sample size" technique. The progressing values of $T$ provides a useful reference framework to help the expert estimate the exponents. This is indicated by the slight pause that the expert usually takes before assessing the exponents for the first decision stage, and the progressive ease with which values for subsequent decision stages are derived.

## LEARNING RESULTS

We performed a case study with the data of about 3801 Duke's Stage C colorectal cancer cases. Beta distributions are used as prior distributions for parameters since all the events are binary. Some of the learned probabilities are shown in Table 1.

Table 1: Prior information and posterior probabilities

| $E(\theta_+)$ | $E(\theta_-)$ | *STAGE* | *STATE* | *EOR* | *EOM* | *TR* | $\alpha_+$ | $\alpha_-$ |
|---|---|---|---|---|---|---|---|---|
| 0.6200 | 0.3800 | 1 | well | N | Y | + | 30 | 70 |
| 0.6010 | 0.3990 | 2 | well | N | Y | + | 60 | 40 |
| 0.3836 | 0.6164 | 3 | well | N | Y | + | 60 | 40 |
| 0.6039 | 0.3961 | 4 | well | N | Y | + | 65 | 35 |
| 0.6041 | 0.3959 | 5 | well | N | Y | + | 70 | 30 |
| 0.6095 | 0.3905 | 6 | well | N | Y | + | 80 | 20 |

## CONCLUSIONS

Compared to its static counterpart, dynamic decision analysis is much more complicated. Since the model parameters vary with time, subjective assessments of the numerical parameters may not be easily and accurately achievable in dynamic decision models. Based on the DynaMoL framework, we have developed a feasible automated probabilities assessment paradigm based on *Bayesian learning, data abstraction*, and *model abstraction*. Preliminary results from a comprehensive case-study in a real-life medical domain have demonstrated the practical promise for this approach.

Acknowledgments

We would like to thank Pheng Kheong Leong for collaborating with us in this work. The work is supported by an Academic Research Grant (No. RP950617) from the National University of Singapore, and a post-doctoral fellowship from the Singapore National Science and Technology Board.